The Materials Project: An application of high-throughput computing

NERSC Brown Bag | May. 2013

Anubhav Jain
Energy & Environmental Technology
Berkeley Lab

Overview

1. The Materials Project

2. High-throughput computing

Materials development is a key bottleneck to realizing renewable energy

solar PV

electric vehicles

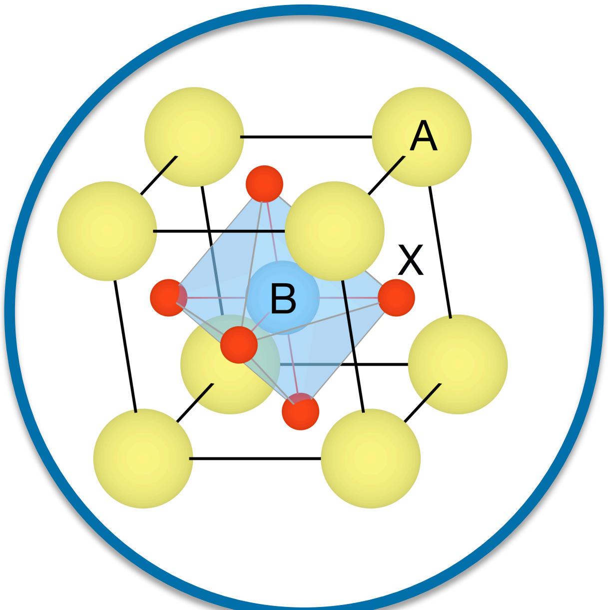
other:

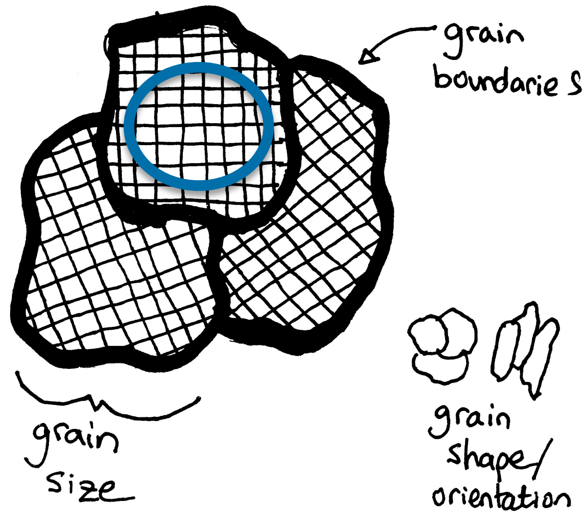
waste heat recovery (thermoelectrics)

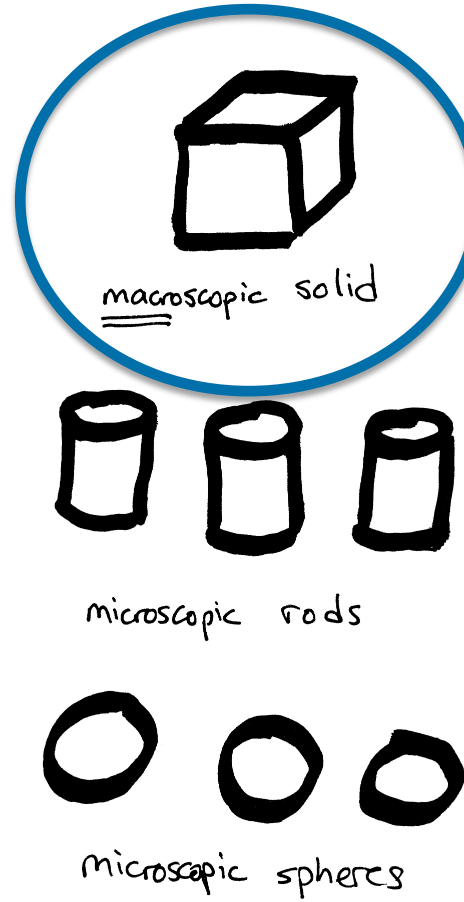
hydrogen storage

catalysts/fuel cells

What do we mean by a material?







Business as usual:
"find the needle in a haystack"

Method one:
Look through the haystack one spot at a time

Hunts Needle in a Haystack

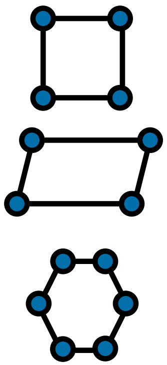
How long does it take to find a needle in a haystack? Jim Moran, Washington, D.C., publicity man, recently dropped a needle into a convenient pile of hay, hopped in after it, and began an intensive search for (a) some publicity and (b) the needle. Having found the former, Moran abandoned the needle hunt.


Found It!!!
Congratulations, it only took you 65298 hours
www.jolyon.co.uk

The Materials Project is like having an army to search through the haystack

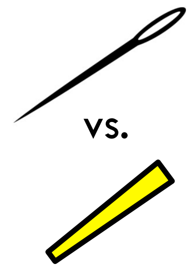
High-throughput computing is like an army

MATERIALS



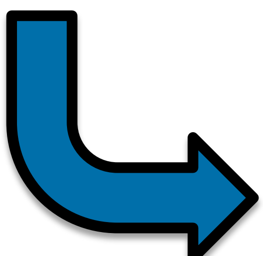
THEORY

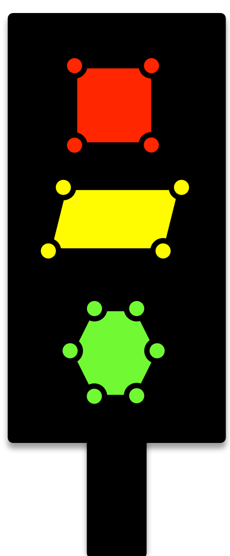
$$i\hbar\frac{d\Psi(\{r_i\};t)}{dt} = \hat{H}\Psi(\{r_i\};t)$$



COMPUTERS

WORKFLOW





Do not synthesize!

Put on backburner

Begin further investigation

Example of a materials problem: Everyone hates batteries

PHONES

Why Your Smartphone Battery Sucks

By Megan Geuss, PCWorld May 18, 2011 6:36 PM

If you're hoping your next smartphone will run faster, shine brighter, connect at 4G speed, *and* last longer on one battery charge, you may be in for a rude surprise. The thirst for battery power in new smartphones and tablets is far outpacing improvements in battery technology. Battery makers are trying to wring the last bits of capacity out of 15-year-old lithium ion technology, while device and app makers seem to be just waking up to the seriousness of the problem. There's an equal share of blame for all parties; meanwhile the immense promise of innovation in mobile devices could come to an early halt due to power limitations, and consumer angst over constantly having to "plug in."

Used to be, you could forget your feature phone's charger at home, go on a long weekend vacation, and--as long as you didn't play hours of Snake--still come home with enough battery life to call a cab. Today, though, we're wedded to our chargers, glaring hawkishly at people who've been hogging airport and coffee shop outlets for too long.

Gadgets

When Lithium-ion Batteries Explode

Paul Mah - May 27, 2008 6:47 PM

Print ShareThis New +1 0 20 comment(s) - last by cogito.. on May 29 at 1:55 AM

Lithium-ion batteries are both a blessing and a curse when it comes to mobile electronics

BlackBerry Curve battery: Cells made in Japan, but assembled where? (Source: Paul Mah)

The topic of exploding lithium-ion batteries has been debated to death in the wake of massive battery recalls over the last couple of years. Amidst the deft public relations maneuvering and finger-pointing, however, the question as to why they explode in the first place is still shrouded in mystery for many.

The most important thing to understand here is that lithium-ion technology is considerably more volatile compared to other forms of rechargeable battery technologies. Defects in the insulating membrane can result in a mini-explosion that rips a battery open to release steam in excess of 600 degrees Fahrenheit.

Manufacturers are aware that it is statistically probable for a lithium-ion to fail, though the calculations employed to sideline the risk are sometimes quite suspect. To determine the mean time between failures (MTBF), manufacturers take a sample of say, 1,000 batteries, which are then used until one fails.

Anatomy of a cathode composition

$$Li_a\, M_b\, (XY_c)_d$$

Li ion source

electron donor / acceptor

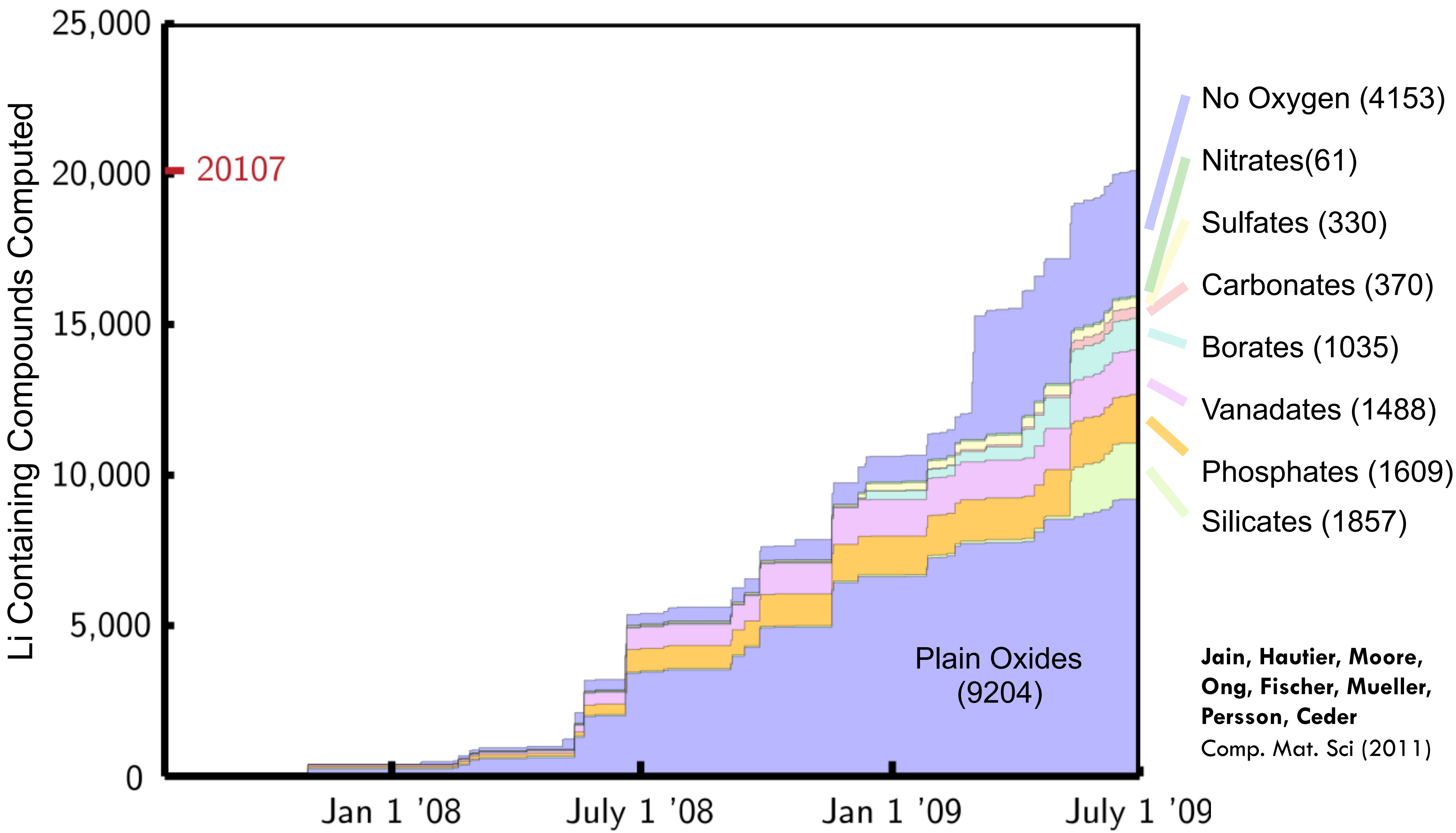
examples: $V^{4+/5+}$, $Fe^{2+/3+}$

structural framework / charge neutrality

examples: O^{2-}, $(PO_4)^{3-}$, $(SiO_4)^{4-}$

common cathodes: $LiCoO_2$, $LiMn_2O_4$, $LiFePO_4$

Compounds screened over time (MIT)



Jain, Hautier, Moore, Ong, Fischer, Mueller, Persson, Ceder
Comp. Mat. Sci (2011)

New mixed phosphate-pyrophosphate material for cathodes

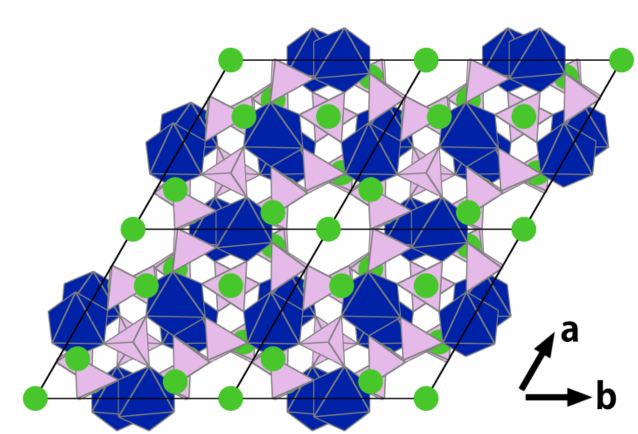
Chemistry	Novelty	Energy density vs. $LiFePO_4$	% of theoretical capacity already achieved in the lab
$Li_9V_3(P_2O_7)_3(PO_4)_2$	New	20% greater	~65%

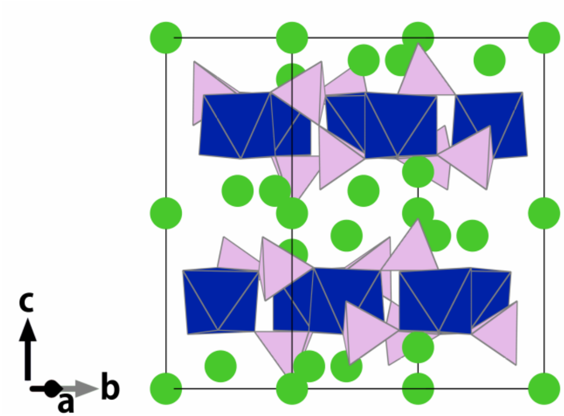
Origin:

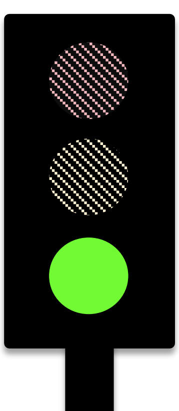
V to Fe substitution in $Li_9Fe_3(P_2O_7)_3(PO_4)_2$*

Remarks:

- Structure has "layers" and "tunnels"
- Pyrophosphate-phosphate mixture
- Potential 2-electron material







Jain, Hautier, Moore, Kang, Lee, Chen, Twu, and Ceder
Journal of The Electrochemical Society 159, A622–A633 (2012).

Discovering a synthesis route

Li_2CO_3
$NH_4H_2PO_4$
V_2O_5

(1) precursors

(2) grind

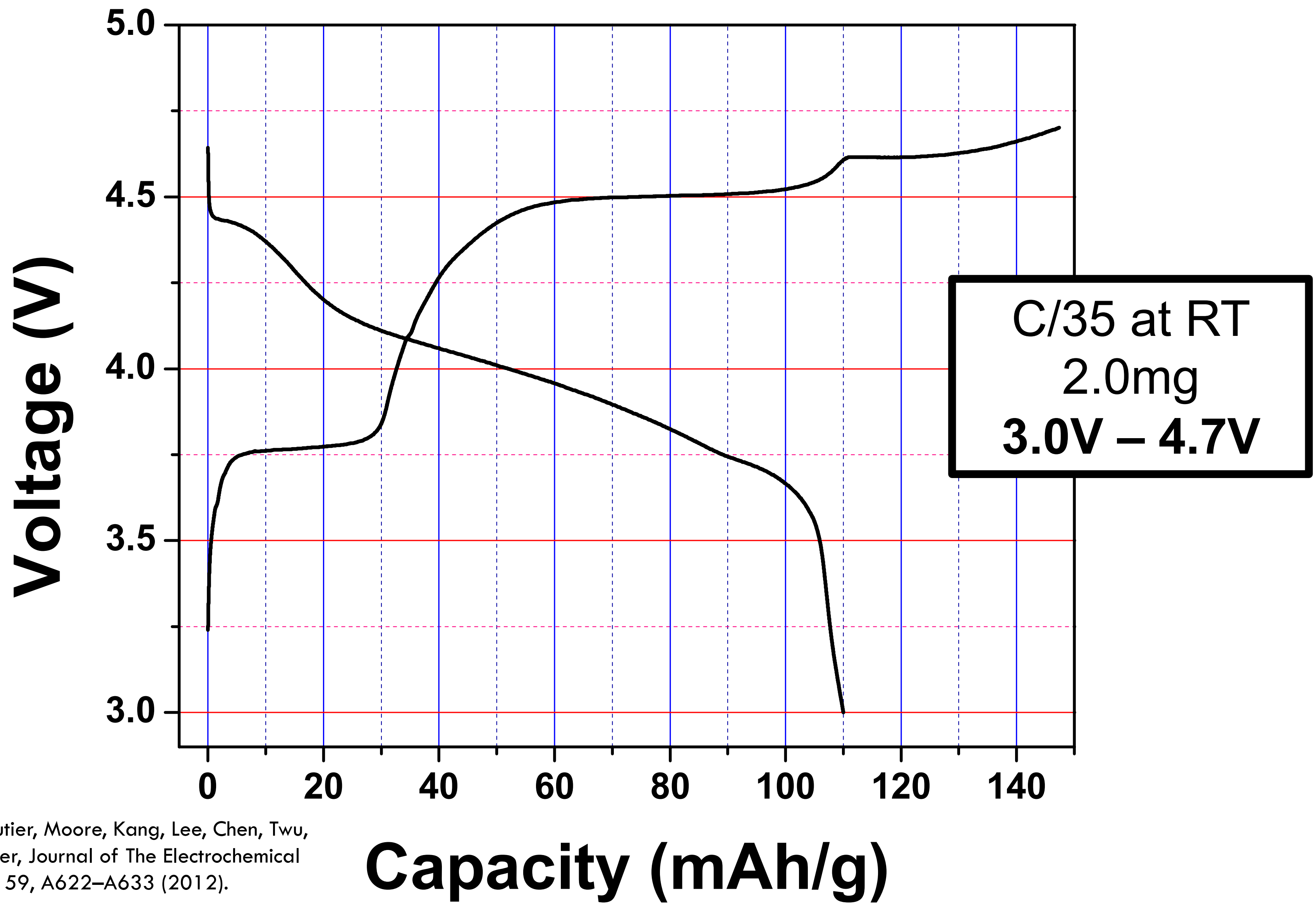
(3) heat 300ºC
6 hours

(4) grind

(5) heat 750ºC
24 hours
97% Ar, 3% H_2

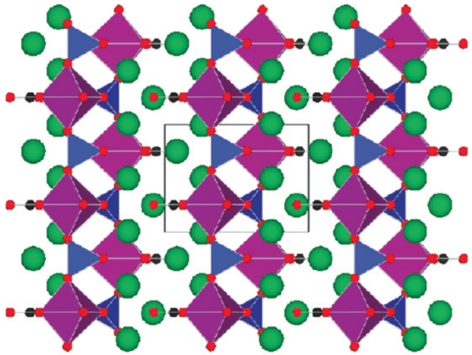
Jain, Hautier, Moore, Kang, Lee, Chen, Twu, and Ceder, Journal of The Electrochemical Society 159, A622–A633 (2012).

Good performance found in initial tests



Jain, Hautier, Moore, Kang, Lee, Chen, Twu, and Ceder, Journal of The Electrochemical Society 159, A622–A633 (2012).

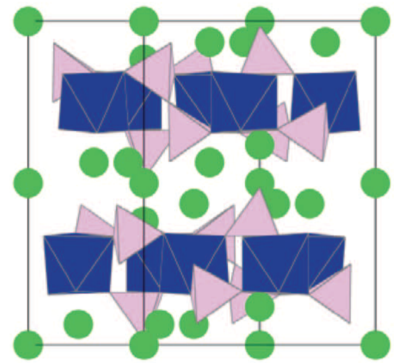
Many new cathodes found for Li ion batteries



Li_3 (Fe, Mn) PO_4 CO_3

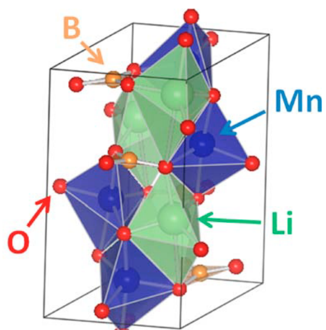
Hautier, Jain, Moore, Chen, Moore, Ong, Ceder
Journal of Materials Chemistry (2011)
Chen, Hautier, Jain, Moore, Kang, Doe et al.
Chemistry of Materials (2012)

(patent filed)



Li_9 V_3 $(P_2O_7)_3$ $(PO_4)_2$

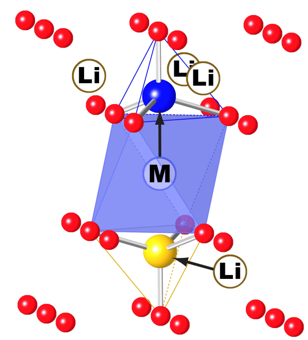
Jain, Hautier, Moore, Kang, Lee, Chen, Twu, Ceder
Journal of the Electrochemical Society (2011)
(patent issued allowance)



Li Mn BO_3

(patent filed)

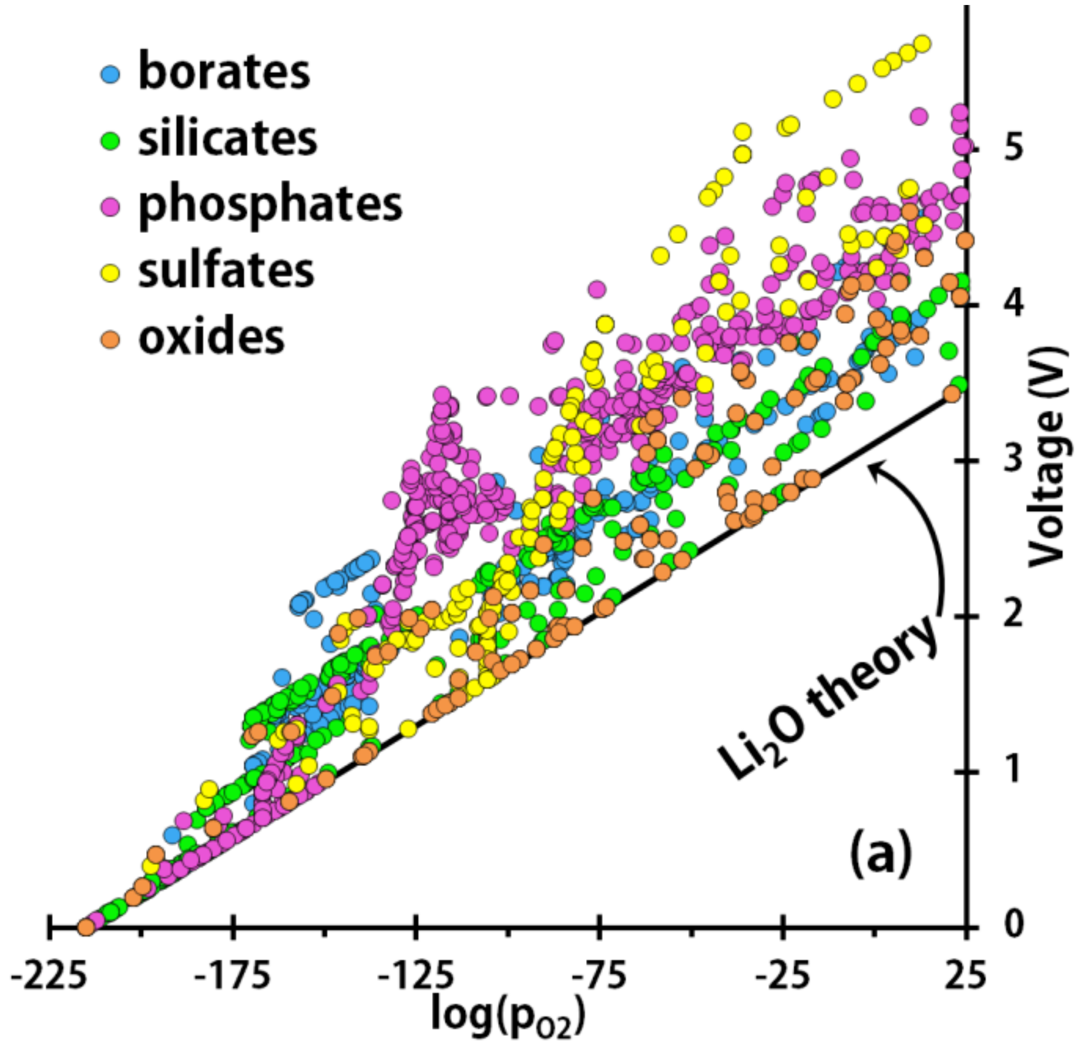
Kim, Moore, Kang, Hautier, Jain, Ceder
Journal of the Electrochemical Society (2011)



Li (V, Cr) O_2

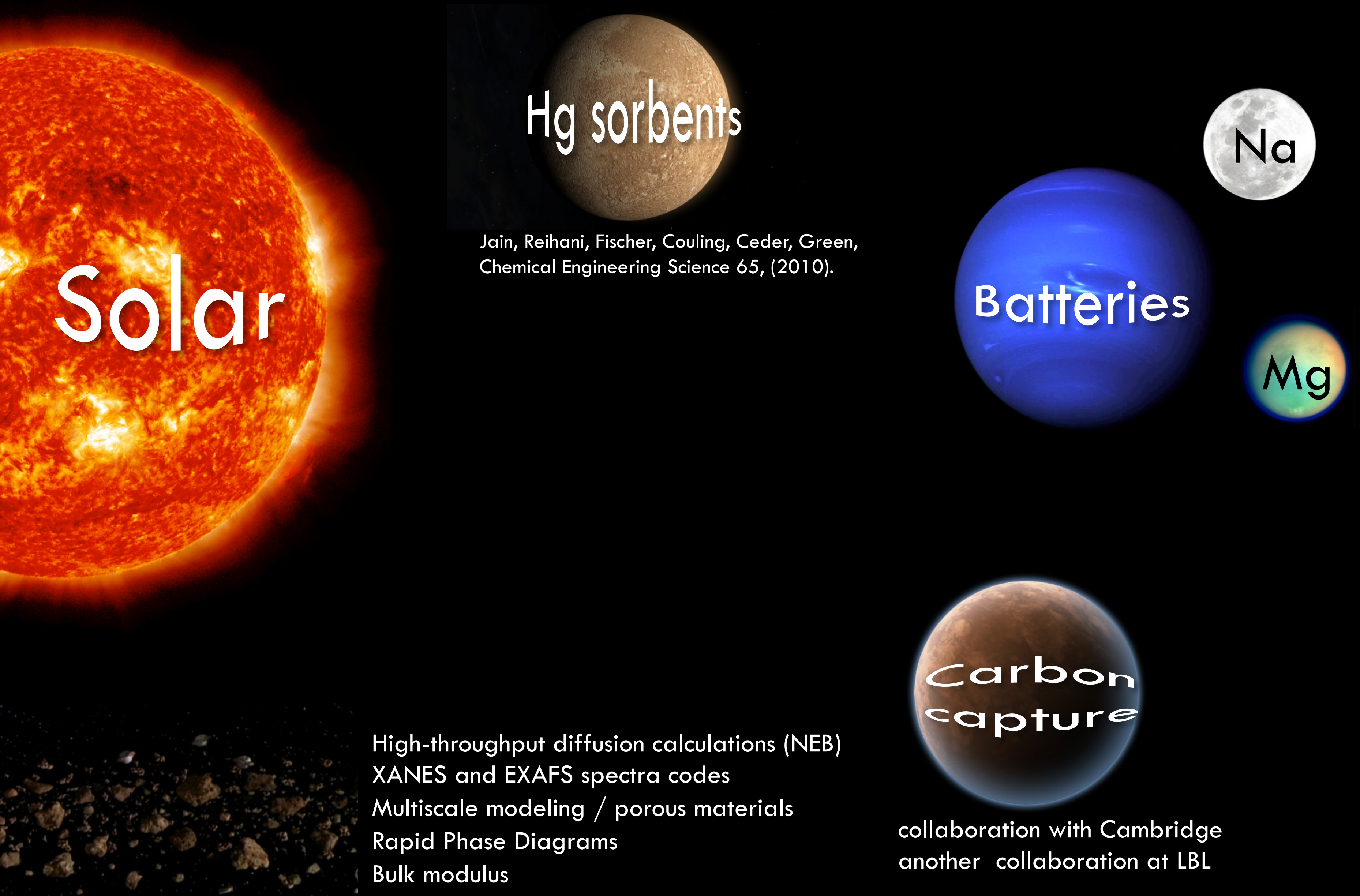
Ma, Hautier, Jain, Ceder
Journal of the Electrochemical Society (submitted)

High voltage materials are less safe: high-throughput finds trends and exceptions



Jain et al.
(in preparation)

Once the army is in place, you can attack more problems



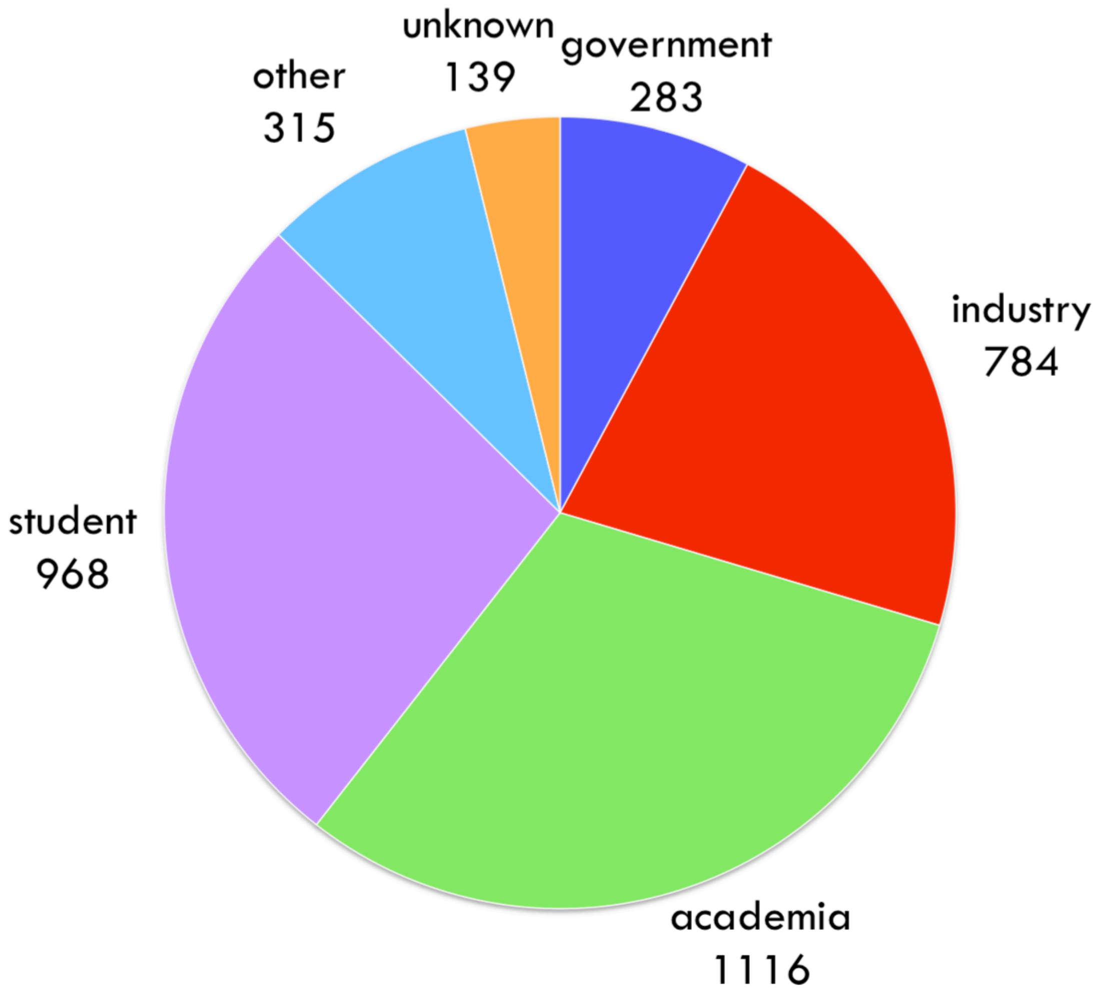
Jain, Reihani, Fischer, Couling, Ceder, Green, Chemical Engineering Science 65, (2010).

High-throughput diffusion calculations (NEB)
XANES and EXAFS spectra codes
Multiscale modeling / porous materials
Rapid Phase Diagrams
Bulk modulus

collaboration with Cambridge
another collaboration at LBL

www.materialsproject.org

3800 users and growing!



(*Old chart for 3600+ users)

User Papers using Materials Project data

Table III. Observed average voltage (vs Li) and the average voltage at 0 K calculated from the Materials Genome Project[15] for some Li_xM alloys.

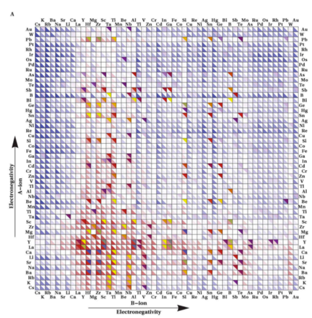
Alloy System	Average Voltage (V) Calculated	Measured
Li_xSi	0.34	0.36
Li_xSn	0.57	0.51
Li_xAl	0.35	0.40

Design of alloy anodes for Li ion batteries

(compared measurements to calculations)

Tran, Obrovac

Journal of the Electrochemical Society (2011)



Computational screening of perovskite metal oxides for optimal solar light capture

(crystal structures for candidate stability screening)

Castelli, Olsen, Datta, Landis, Dahl, Thygesen, Jacobsen

Energy & Environmental Science (2011)

TABLE 16. Energies calculated from the MGP Reaction Calculator vs. measured enthalpies taken from MGP database, and the ratios of these numbers

	$\Delta E^{f}_{ox}(0)$ kJ/mol	ΔH^{f}_{ox} kJ/mol	$\Delta E^{f}_{ox}(0)/\Delta H^{f}_{ox}$
$MgCO_3$	–81	–117	0.69
$CaCO_3$	–148	–178	0.83
$SrCO_3$	–198	–235	0.84
$BaCO_3$	–238	–276	0.86

DFT Formation Enthalpies of Rare Earth Orthophosphates

(find scaling factor for computed heat of formation)

Rustad

American Minerologist (2012)

Table 2
Most favora... ases and their reaction energies according to the *Materials Project* database. There are no stable binary phases in the Cr–Fe–Sn system.

Reaction	ΔE_0 (eV/f.u.)
$3(2Cr + Fe + Si) \rightarrow 2Cr_3Si + Fe_3Si$	–1.35
$3(2Cr + Fe + Ge) \rightarrow 2Cr_3Ge + Fe_3Ge$	–0.45
$2Cr + Fe + Sn \rightarrow 2Cr + Fe + Sn$	0.00
$2Cr + Co + Al \rightarrow 2Cr + CoAl$	–1.20
$2Cr + Co + Ga \rightarrow 2Cr + CoGa$	–0.56
$3(2Cr + Co + In) \rightarrow In_3Co + 6Cr + 2Co$	–0.05

Phase stability of chromium based compensated ferrimagnets with inverse Heusler structure

(find reaction paths and energies)

Meinert, Geisler

Journal of Magnetism and Magnetic Materials (2013)

"Thanks. Your product is astounding. I redid work for a recent paper that took weeks in about 15 minutes! I guess this is truly “transformative” science in the NSF sense!"

What's next for Materials Project?

- More compounds, more properties, more apps
- Increased adoption and support of REST interface and programmatic access to data
- Calculations on Demand
- Sandboxes

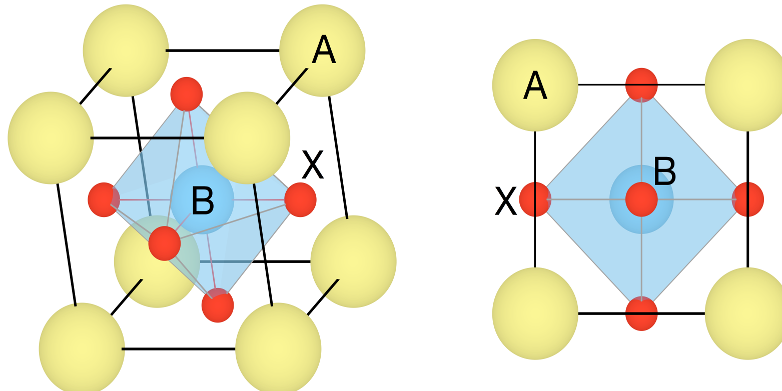
Can we do better than an army?

Recap method one:
Look through the haystack one spot at a time

Recap method two:
Hire an army to search through the haystack

Method three:
Equip your army with metal detectors

Searching for perovskite water splitters



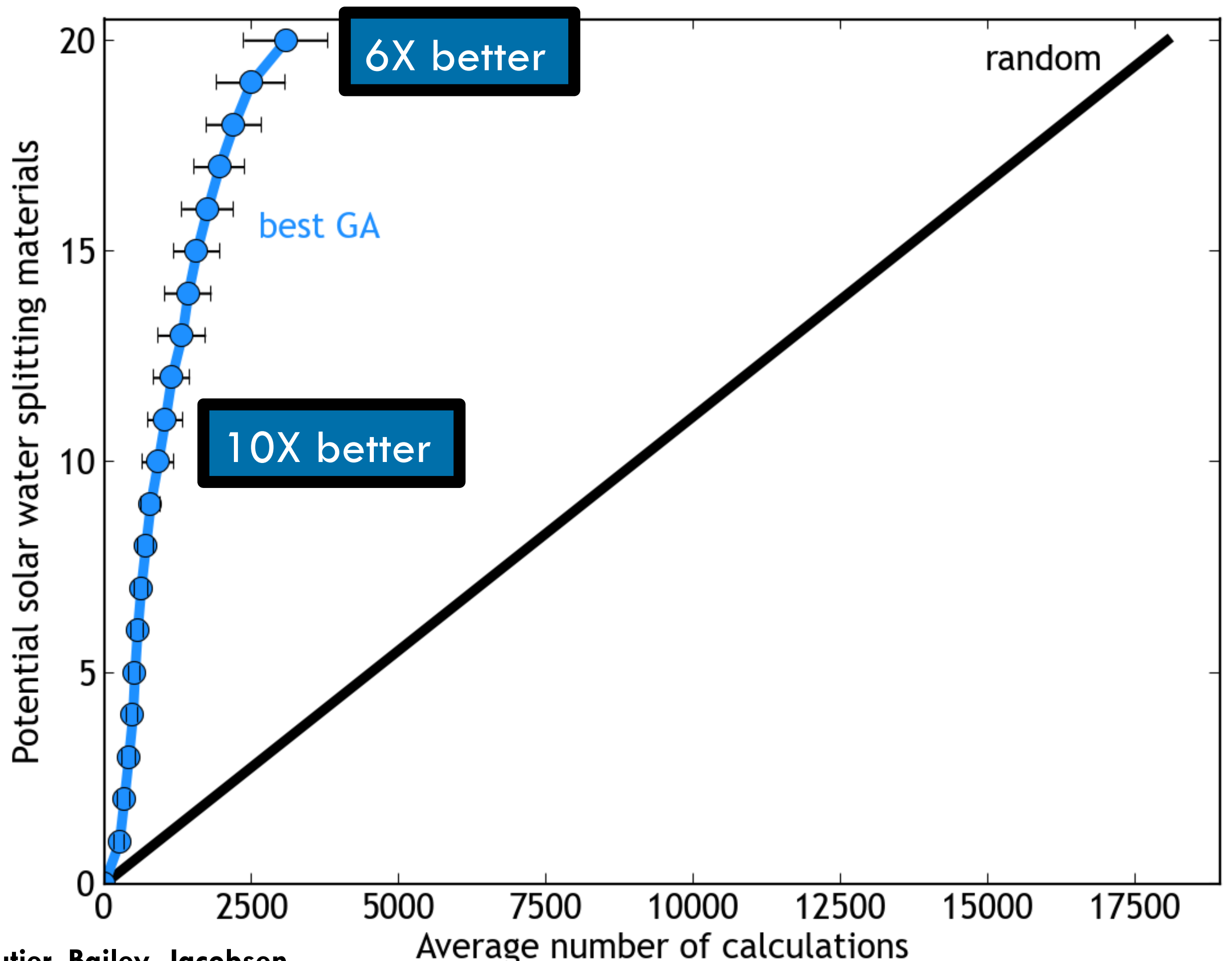
A B X_3

52 metals | 52 metals | 7 mixtures {O, N, F, S}

(about 19,000 total compounds!)

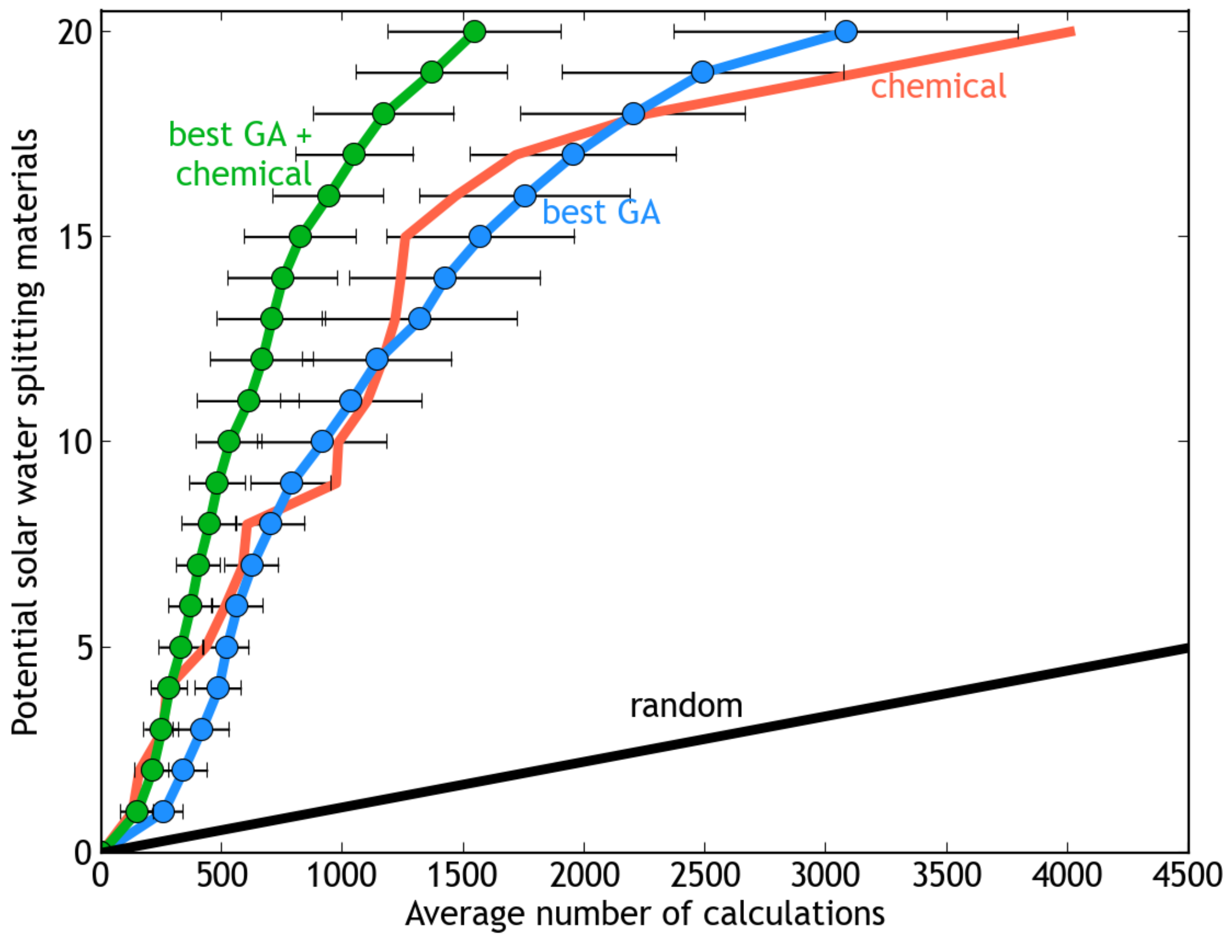
examples: $SnTiO_3$, $SrGeO_3$

Without any prior knowledge, a GA can be 5 – 10 times better than random search



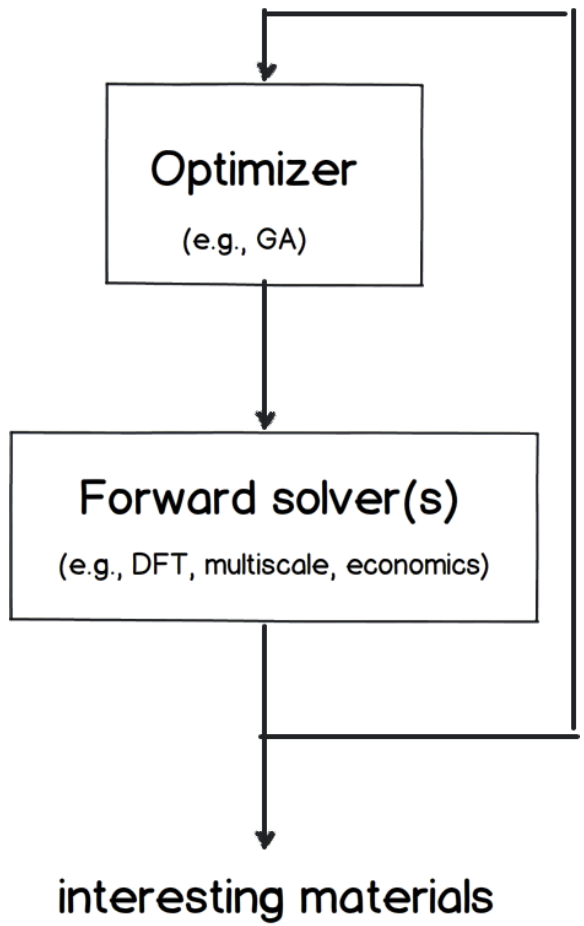
Jain, Castelli, Hautier, Bailey, Jacobsen
Accepted, J. Materials Science 2013

An uninformed GA is about as good as basic chemical rules



Jain, Castelli, Hautier, Bailey, Jacobsen
Accepted, J. Materials Science 2013

Could we suggest materials automatically?
Develop a forward solver, press a button…

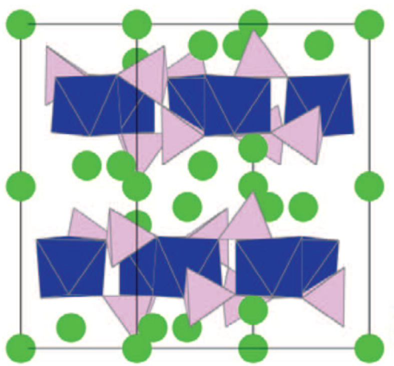


Overview

1. The Materials Project

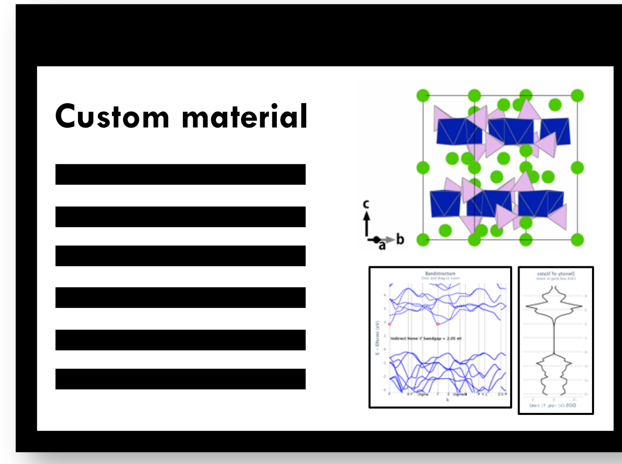
2. High-throughput computing

The high-throughput workflow drives you from materials to properties



A cool material !!

WORKFLOW



Lots of information about cool material !!

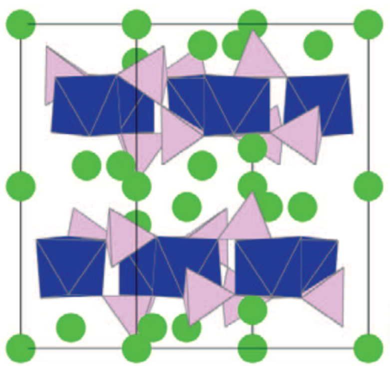
$$i\hbar \frac{d\Psi(\{r_i\};t)}{dt} = \hat{H}\,\Psi(\{r_i\};t)$$

Physicist/ Scientist

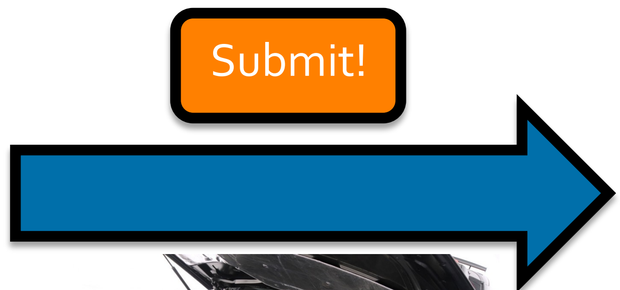
#*&$ PBS!!

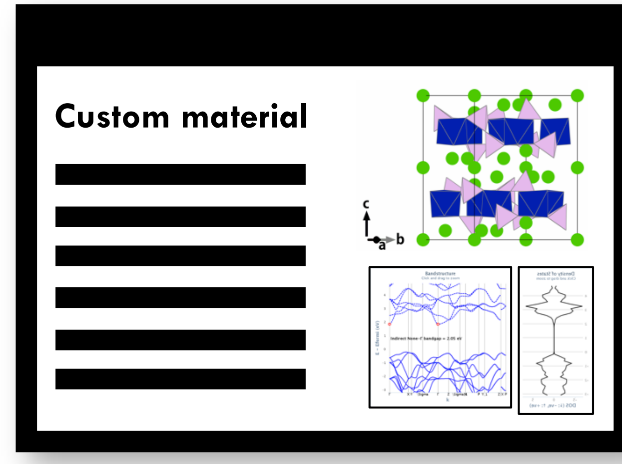
"Software Engineer"

The workflow takes care of a lot "under the hood"

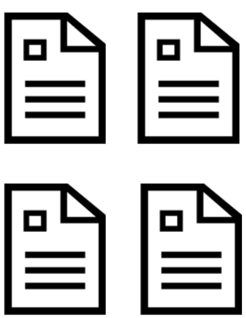


A cool material !!

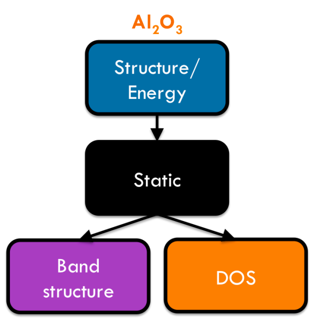




Lots of information about cool material !!

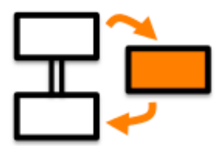


Input generation (parameter choice)

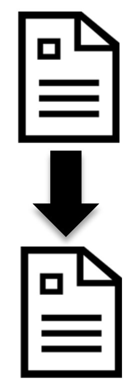


Workflow mapping

Supercomputer submission / monitoring



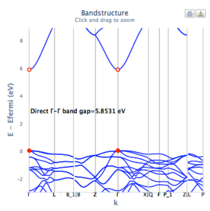
Error handling

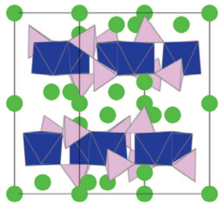


File Transfer

File Parsing / DB insertion

Infrastructure snapshot



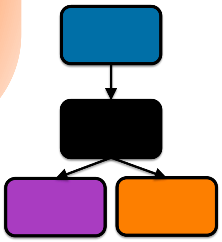


Materials Data

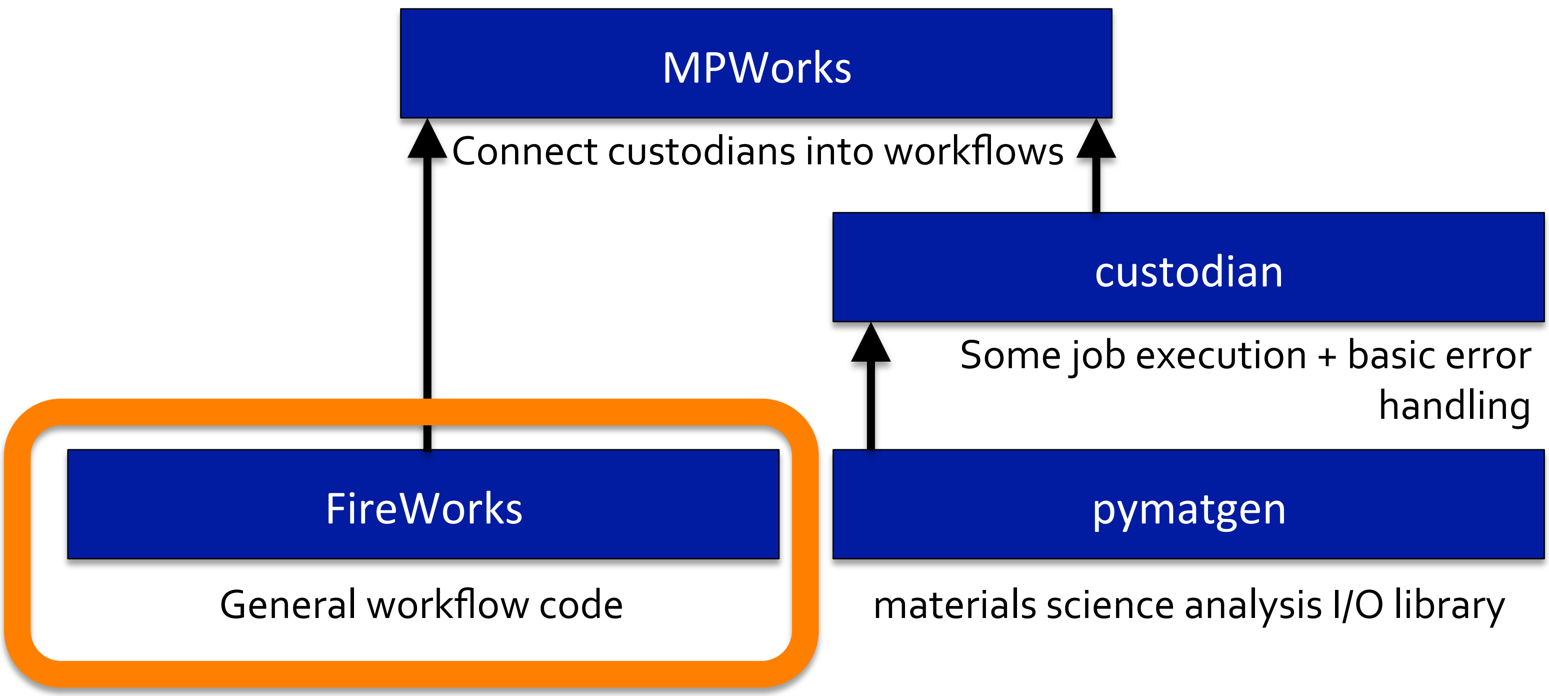
- Over 10 million CPU hours of calculations in < 6 months
- Over 40,000 successful VASP runs (30,000+ materials)
- Generalizable to other high-throughput codes/analyses

Submitted Materials

Calculation Workflows



Overview of our (new) codebases



python

The key ingredients of FireWorks

LaunchPad (Adder/Monitor)

Adds jobs (FireWorks) to database

Monitor status of FireWorks

Python script

Rocket Launcher (Runner)

Pulls a FireWork from the database

Runs it

Updates FireWorks database

Python script

FireWork (Job)

Represents a computing job

Contains all information needed to run a job

Stored in MongoDB

How does it work (command line)

- Pretend you have a computing job defined as “**FW1.yaml**” which is an instruction to write some text into a file
  - We’ll cover this file later
  - It encapsulates everything you need to know in order to bootstrap a job
    - Script to run
    - Parameters to run with (e.g., pass a dict into Python code)

How it works

```
>> lpad add fw1.yaml
>> rlaunch singleshot
```

Variations on a theme

- So far, we have two basic functions
  - Add a job to a database
    - (lpad add)
  - Pull a job, run it, and update database
    - (rlaunch singleshot)

- How can we get more complex functions?

Running on a head node

```
>> rlaunch singleshot
```

(we already saw this)

Running on a queue

```
>> qlaunch singleshot
```

- Submit job that runs rlaunch

```
#PBS -l nodes=1:ppn=8
#PBS -l walltime=00:30:00

cd <my_dir>
rlaunch singleshot
```

- Will grab a job from the database and run it
- All your queue scripts can be identical

Running and queuing multiple jobs

- To run many jobs, use rapidfire:
  - `rlaunch rapidfire`
  - `qlaunch rapidfire`
- Options for:
  - how many jobs to run (or infinite)
  - how many jobs to maintain in queue
  - Run once or “sleep and reprocess” in case new jobs added to DB

More complex?

- Head node
  - rlaunch singleshot

PBS, SGE, SLURM

- Queue
  - Submit “dummy” script
  - To run multiple FireWorks in 1 PBS job, replace `rlaunch singleshot` with `rlaunch rapidfire`

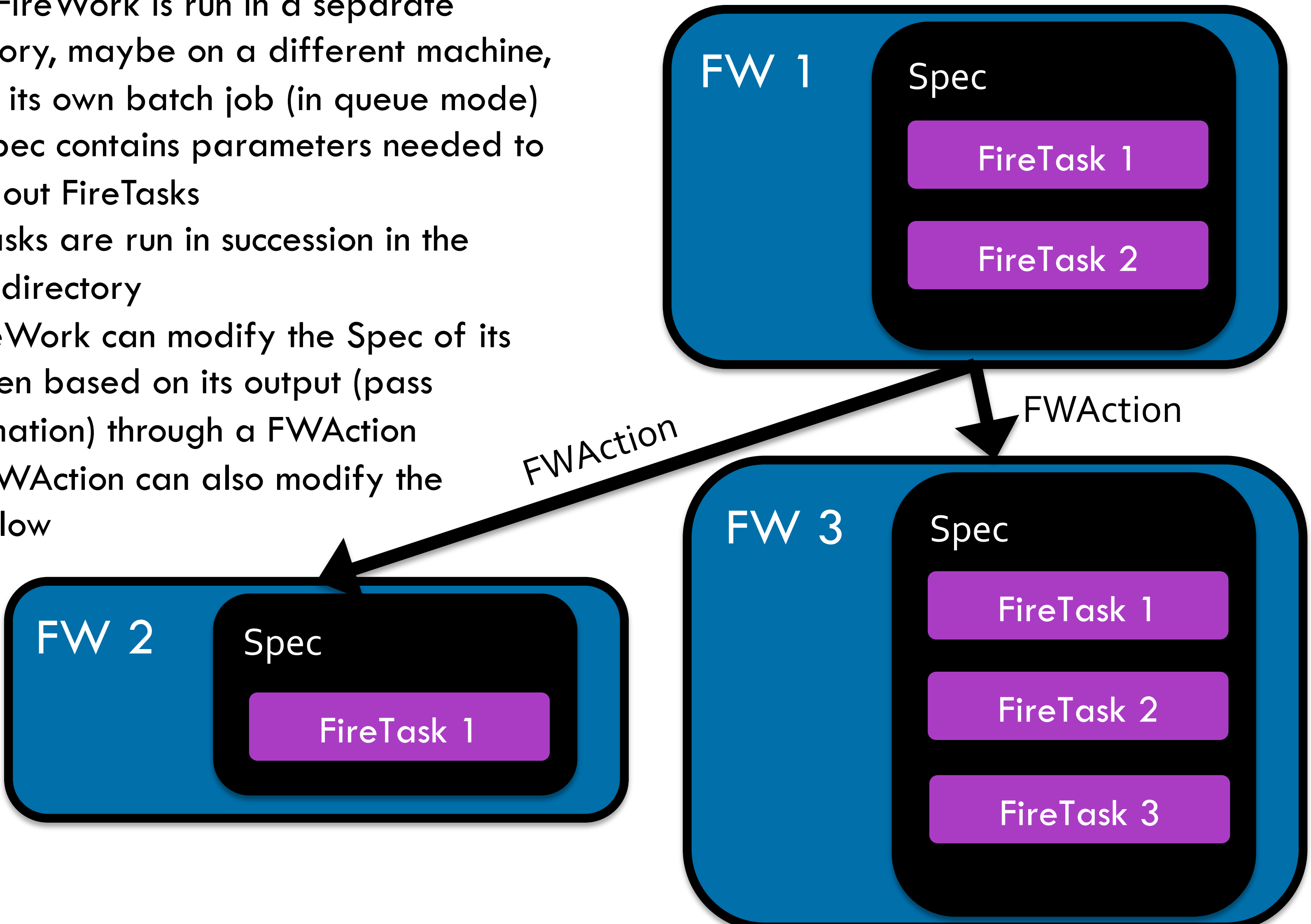
- Simple Job Packing (?)
  - run a thin MPI wrapper that submits **rlaunch rapidfire** to several nodes at once?

newt

- NEWT(?)
  - Submit `rlaunch` jobs through API?

FireWorks also have variations

- Each FireWork is run in a separate directory, maybe on a different machine, within its own batch job (in queue mode)
- The spec contains parameters needed to carry out FireTasks
- FireTasks are run in succession in the same directory
- A FireWork can modify the Spec of its children based on its output (pass information) through a FWAction
- The FWAction can also modify the workflow



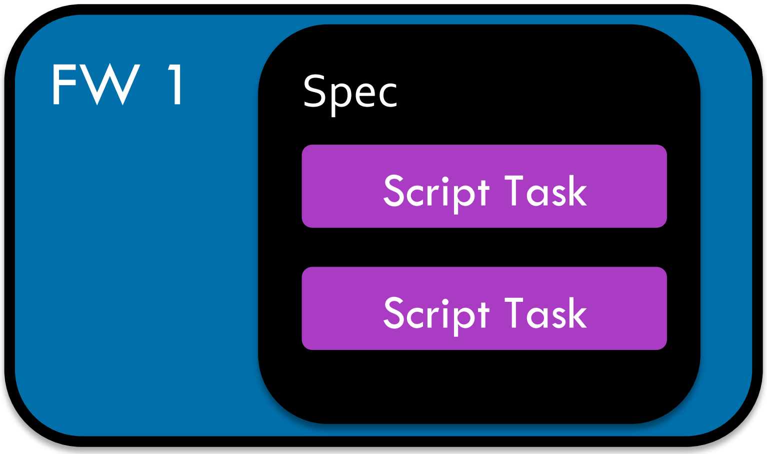
Example 1 – single FW, single task

A FireTask with _fw_name = "Script Task" is in:
fireworks/user_objects/firetasks/script_task.py

fw_tutorials/installation/fw_test.yaml

```
spec:
  _tasks:
 - _fw_name: Script Task
    script: echo "howdy, partner!>> howdy.txt
```

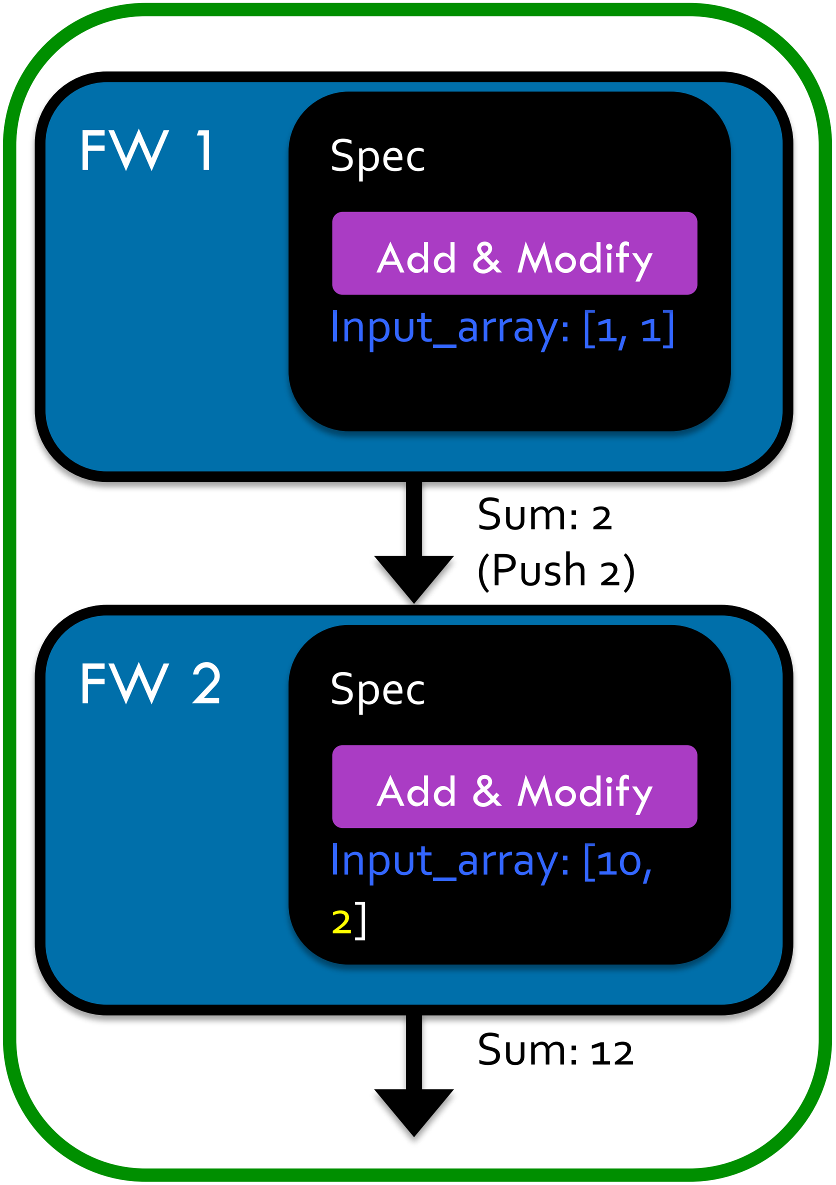
Example 2 – single FW, multi task



fw_tutorials/firetask/fw_multi.yaml

```
spec:
  _tasks:
  - _fw_name: Script Task
    script: echo "howdy, partner!" > howdy.txt
  - _fw_name: Script Task
    script: wc -w < howdy.txt > words.txt
```

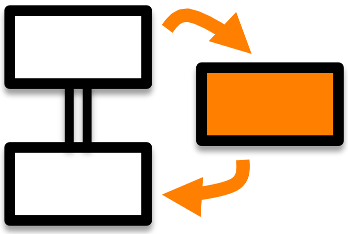
Example 3 – information passing



fw_tutorials/dynamic_wf/addmod_wf.yaml

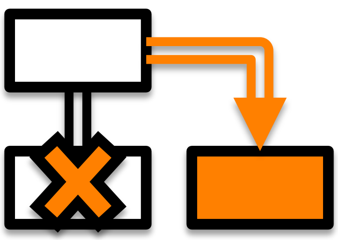
```yaml
fws:
- fw_id: -1
  spec:
    _tasks:
    - _fw_name: Add and Modify Task
    input_array:
    - 1
    - 1
- fw_id: -2
  spec:
    _tasks:
    - _fw_name: Add and Modify Task
    input_array:
    - 10
links:
  -1:
  - -2
metadata: {}
```

Dynamic workflows



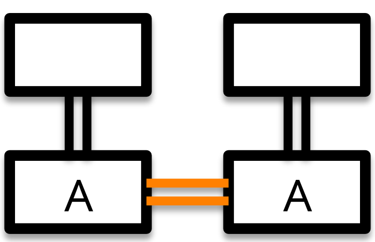
Detours

(about 10-20% of jobs fail and must be rerun with different input parameters)



Branches

(based on the result of a calculation, the entire workflow might need to be modified)



Duplicate Job detection

(if two workflows contain an identical step, ensure that the step is only run once and relevant information is still passed)

Automated Workflows with FireWorks

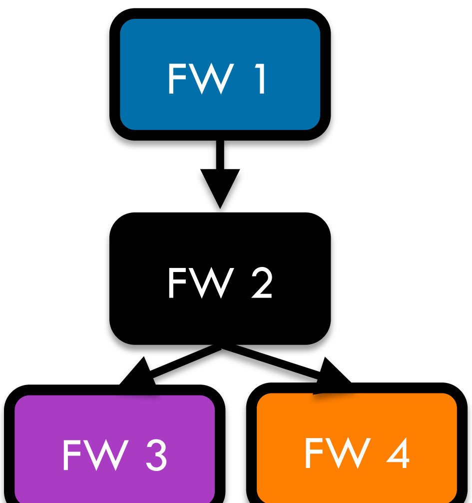
ROCKET LAUNCHER / QUEUE LAUNCHER

Directory 1

Directory 2

Directory 3

Directory 4



LAUNCHPAD

A lot more is possible…

- Assigning different walltimes/ncores to different types of jobs
- Directing certain calculations to certain machines
- Priorities
- Failure handling / rerunning jobs
- Tracking jobs (running/completed/etc) and performing complicated queries
- More…

Limitations

- Not been stress-tested to hundreds of jobs within a single workflow.
- Not been stress-tested to millions of workflows.
- Doesn’t automatically optimize the distribution of computing tasks over worker nodes

FireWorks is open source

(http://pythonhosted.org/FireWorks)

- Installation is simple for most users (Windows?):
  - Prereqs: Mongo, pip, git
- Step-by-step tutorials cover everything
- I am happy help you get started

What are the stumbling blocks?

- **rlaunch** requires you to query and update an external database
  - Firewall issue in accessing external db?
  - We only got things working on Hopper/Carver by having NERSC host a database on their network
- **rlaunch rapidfire --nlaunches infinite** usually means persistent script running on head node
  - NERSC is very strict about kicking off persistent scripts
  - Script is lightweight but "always there"
  - Suffice to say our solution is not ideal…

Conclusion / Final thoughts

- Materials Project is an example of ‘Science Gateway’ based on HT-computed data
- At NERSC, viable options for HT already exist
  - We use thruput queue heavily
  - new tools under development (NEWT)
- Different groups are developing solutions
  - Will be interesting to know what ‘sticks’, or if equilibrium will be unique workflow per project
  - Would love to discuss it with you!

Thank you!

- Materials Project
  - Kristin Persson, Gerbrand Ceder, Shyue Ping Ong, Geoffroy Hautier, Michael Kocher, David Skinner, David Bailey, Dan Gunter, Shreyas Cholia, NERSC staff
- Inverse Design
  - Ivano Castelli, Geoffroy Hautier, David Bailey, Karsten Jacobsen
- Funding
  - DOE, NSF, 2010 LBL LDRD
- Contact:
  - ajain@lbl.gov